

Bagan Kendali *Robust* Multivariat untuk Pengamatan Individual

*Seltuti, Erna Tri Herdiani^{*1}, Nirwan Ilyas^{*2}

Abstract

The most widely used of control chart in multivariate control processing is control chart T^2 Hotelling. There are 2 kinds of control chart T^2 Hotelling, namely T^2 Hotelling for group observation and T^2 Hotelling for individual observation. In this paper, discuss the control chart T^2 Hotelling for individual observation. This control chart is used for monitoring of mean vector and sample of covariance matrix. Mean vector and sample of covariance matrix are very sensitive with respect to extreme point (*outliers*). Therefore, it is needed an estimator of mean vector and has a stocky population covariance matrix to the outliers data. One method that can be used to detect data that contains outliers is *Minimum Covariance Determinant* (MCD). From the calculation results, obtained that control chart T^2 Hotelling by using *Fast-MCD* algorithm is more sensitive to detect outliers data than T^2 Hotelling classically.

Keyword: T^2 Hotelling, *Minimum Covariance Determinant* (MCD), *robust*, *outlier*

Abstrak

Bagan kendali yang paling banyak digunakan dalam pengendalian proses secara multivariat adalah bagan kendali T^2 Hotelling. Ada 2 jenis dari bagan kendali T^2 Hotelling yaitu bagan kendali T^2 Hotelling untuk pengamatan kelompok dan individual. Pada tulisan ini membahas bagan kendali T^2 Hotelling untuk pengamatan individual. Bagan kendali ini digunakan untuk memonitor vektor rata-rata dan matriks kovariansi sampel. Vektor rata-rata dan matriks kovariansi sampel sangat sensitif terhadap titik ekstrim (*outliers*). Oleh karena itu dibutuhkan estimator vektor rata-rata dan matriks kovariansi populasi yang kekar terhadap data *outliers*. Salah satu metode yang dapat digunakan untuk mendeteksi data yang mengandung *outliers* adalah *Minimum Covariance Determinant* (MCD). Dari hasil perhitungan diperoleh bahwa bagan kendali T^2 Hotelling dengan algoritma *Fast-MCD* lebih sensitif mendeteksi data *outliers* daripada T^2 Hotelling klasik.

Kata Kunci: T^2 Hotelling, *Minimum Covariance Determinant* (MCD), *robust*, *outlier*.

1. Pendahuluan

Bagan kendali kualitas atau yang disebut *control chart* merupakan salah satu alat yang digunakan dalam usaha mengendalikan kualitas proses karena dalam bagan kendali dapat diketahui kapan proses berada di luar kendali (*out of control*). Untuk mengukur dua atau lebih karakteristik kualitas secara bersamaan digunakan bagan kendali multivariat yang diperkenalkan oleh Harold Hotelling. Bagan kendali ini dikenal sebagai bagan kendali T^2 Hotelling. Menurut Montgomery (2009), bagan kendali T^2 Hotelling paling banyak digunakan dalam pengendalian proses secara multivariat untuk memonitor vektor rata-rata proses karena dalam bagan kendali T^2 Hotelling menggunakan vektor rata-rata dan matriks kovariansi dari sampel. Vektor rata-rata dan matriks kovariansi sampel sangat sensitif terhadap titik ekstrim (*outliers*). Karena itu dibutuhkan estimator vektor rata-rata dan matriks kovariansi populasi yang kekar terhadap data penciran untuk membuat bagan kendali T^2 Hotelling.

Salah satu metode yang cukup terkenal adalah *Minimum Covariance Determinant* (MCD)[9]. Menurut berbagai penelitian antara lain Rousseeuw dan van Driessen (1998), Hubert dan van Driessen

*Program Studi Statistika, Departemen Matematika, FMIPA, Universitas Hasanuddin

¹herdiani.erna@gmail.com, ²nirwanilyas@yahoo.com

(2002), penaksir *robust* dengan metode MCD dapat memberikan angka ketepatan yang lebih baik dibanding metode-metode yang lainnya. Namun, pada metode MCD, apabila data berdimensi cukup besar, akan sulit dilakukan karena terdapat banyak kombinasi sub sampel yang harus ditemukan oleh penaksir MCD. Oleh karena itu, Roussew dan Van Driessen (1999) menemukan algoritma *Fast-MCD*. Dari sudut pandang ketersediaan paket program, MCD telah terakomodir dalam S-PLUS, Matlab dan SAS dengan menggunakan algoritma *Fast-MCD*. Penaksir *Fast-MCD* merupakan estimasi dengan menggunakan rata-rata dan kovariansi dari sebagian pengamatan yang meminimumkan determinan matriks kovariansi. Metode *Fast-MCD* selain memenuhi sifat statistik *high breakdown point* (dapat mengatasi pencilan sampai 50%), juga efisien dalam komputasinya dan yang paling penting adalah efektif dalam mengatasi pengamatan yang mengandung pencilan. Berdasarkan hal tersebut, dalam penulisan tugas akhir ini membahas bagan kendali *robust* multivariat untuk pengamatan individual diterapkan untuk menentukan pengamatan yang *out of control* dari bahan Clay Putih sebagai bahan pengisi dan pengikat untuk pupuk NPK Phonska.

2. Tinjauan Pustaka

2.1 Pengendalian Kualitas

Pengendalian proses multivariat merupakan salah satu bagian dari pengendalian kualitas yang cepat berkembang karena ada banyak situasi *real* yang melibatkan lebih dari dua karakteristik kualitas proses yang saling berhubungan. Pengendalian proses multivariat ini selanjutnya dikenal sebagai *Multivariate Statistic Proses Control* (Erna, 2014). Seperti halnya *SPC*, dalam *MSPC* juga terdapat bagan kendali. Bagan kendali multivariat digunakan untuk memantau peubah secara bersama-sama pada suatu proses. Proses yang dilakukan dikatakan terkendali jika tidak ada pengamatan yang berada di luar batas kendali (*outlier*).

2.2 Analisis Multivariat

Analisis multivariat merupakan analisis yang melibatkan banyak variabel (lebih dari dua). Asumsi-asumsi pada analisis multivariat adalah variabel karakteristik harus saling berkorelasi dan berdistribusi normal multivariat [7].

2.2.1 Uji Korelasi

Korelasi adalah suatu ukuran yang menyatakan kekuatan hubungan antara 2 variabel [9]. Perhitungan koefisien korelasi Pearson antara variabel X_j dan X_k , yaitu :

$$\begin{aligned} \rho_{X_j X_k} &= \frac{\text{cov}(X_j, X_k)}{\sqrt{\text{var}(X_j, X_j)} \sqrt{\text{var}(X_k, X_k)}} \\ &= \frac{\sum_{i=1}^N (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)}{\sqrt{\sum_{i=1}^N (X_{ij} - \bar{X}_j)^2} \sqrt{\sum_{i=1}^N (X_{ik} - \bar{X}_k)^2}} \end{aligned}$$

Berikut merupakan uji korelasi.

1. Hipotesis :

$H_0: \rho = \mathbf{I}$ (tidak terdapat korelasi antar variabel).

$H_1: \rho \neq \mathbf{I}$ (terdapat korelasi antar variabel).

2. Statistik uji :

$$T = \frac{(n-1)}{(1-r)^2} [\sum \sum_{j < k} (r_{jk} - \bar{r})^2 - \hat{\gamma} \sum_{k=1}^p (\bar{r}_k - \bar{r})^2] \quad (1)$$

dengan :

$$r_{jk} = r_{X_j X_k}$$

$$\begin{aligned}\bar{r}_k &= \frac{1}{p-1} \sum_{\substack{j=1 \\ j \neq k}}^p r_{jk} ; k = 1, 2, \dots, p \\ \bar{r} &= \frac{2}{p(p-1)} \sum \sum_{j < k} r_{jk} \\ \hat{\gamma} &= \frac{(p-1)^2 [1 - (1 - \bar{r})^2]}{p - (p-2)(1 - \bar{r})^2}\end{aligned}$$

3. Kriteria Pengujian :

Jika statistik uji $T > \chi^2_{\frac{(p+1)(p-2)}{2}, (\alpha)}$, maka H_0 ditolak. Sehingga dapat disimpulkan H_1 diterima bahwa terdapat korelasi yang signifikan antar variabel.

2.2.2 Distribusi Normal Multivariat

Pada bagan kendali multivariat, pengujian distribusi normal multivariat dilakukan untuk menguji apakah distribusi data yang akan dianalisis telah berdistribusi normal multivariat sebagai asumsi dasar yang harus dipenuhi dalam analisis multivariat [7].

Berikut pengujian distribusi normal multivariat:

1. Hipotesis :

H_0 = data berdistribusi normal multivariat.

H_1 = data tidak berdistribusi normal multivariat.

2. Statistik Uji :

$$\begin{aligned}d_i^2 &= (\mathbf{X}_{ij} - \bar{\mathbf{X}}_j)^T \mathbf{S}^{-1} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_j); \\ i &= 1, 2, \dots, n; j = 1, 2, \dots, p\end{aligned} \quad (2)$$

3. Kriteria pengujian

Data dapat dikatakan berdistribusi normal multivariat atau dapat dikatakan gagal menolak H_0 apabila terdapat minimal 50 % nilai $d_i^2 \leq \chi^2_{(p, 0.5)}$ atau hasil dari *scatter plot* berupa garis lurus [7].

2.3 Estimator Robust Minimum Covariance Determinant (MCD)

Menurut Hubert dkk (2005) penaksir *robust Minimum Covariance Determinant* (MCD) merupakan rata-rata kovariansi dari sebagian pengamatan yang meminimumkan determinan matriks kovariansi. Misalkan vektor random $\mathbf{X} = [X_1, X_2, \dots, X_p]^T$ merupakan himpunan data sejumlah n pengamatan yang terdiri dari p variabel, dimana $n \geq p + 1$. Penaksir MCD merupakan pasangan rata-rata sub sampel $\bar{\mathbf{x}}$ dan \mathbf{S} yang merupakan matriks definit positif simetri berdimensi $p \times p$ dari suatu sub sampel berukuran h pengamatan dimana $\frac{n+p+1}{2} \leq h \leq n$ dengan

$$\bar{\mathbf{x}} = \frac{1}{h} \sum_{i=1}^h x_i = \frac{x_1 + x_2 + \dots + x_h}{h} \quad (3)$$

$$\mathbf{S} = \frac{1}{h-1} \sum_{i=1}^h (x_i - \bar{\mathbf{x}})(x_i - \bar{\mathbf{x}})' \quad (4)$$

Algoritma yang terkenal dalam menaksir estimator MCD adalah *Fast-MCD* yang diusulkan oleh Rousseeuw dan van Driessen (1999).

Algoritma *Fast-MCD* dapat dilakukan dengan langkah-langkah sebagai berikut:

1. Ambil sejumlah h pengamatan yang berbeda secara acak. Dari n pengamatan akan dihasilkan $\binom{n}{h}$ himpunan yang baru. Nilai h yang optimal memenuhi $(n + p + 1)/2$.

2. Definisikan himpunan pertama sebagai H_1 . Berdasarkan himpunan H_1 hitung vektor rata-rata dan matriks kovariansi (\bar{x}_1, S_1) menggunakan rumus pada persamaan (3) dan (4).
3. Menghitung jarak mahalnobis dengan menggunakan rumus:

$$d_1(i) = \sqrt{(x_i - \bar{x}_1)' S_1^{-1} (x_i - \bar{x}_1)}$$

4. Mengurutkan $d_1(i)$ dari nilai terkecil ke nilai yang terbesar.
5. Definisikan himpunan bagian baru dengan H_2 , sedemikian sehingga $\{d_1(i); i \in H_2\} := \{(d_1)_{1:n}, (d_1)_{2:n}, \dots, (d_1)_{h:n}\}$, dimana $(d_1)_{1:n} \leq (d_1)_{2:n} \leq \dots (d_1)_{h:n}$.
6. Menghitung vektor rata-rata dan matriks kovariansi (\bar{x}_2, S_2) dari H_2
7. Bandingkan $\det(S_2)$ dengan $\det(S_1)$. Bila $\det(S_2) > \det(S_1)$ ulangi langkah pada poin (1) sampai (6) hingga ditemukan bahwa $\det(S_{m+1}) \leq \det(S_m)$.

2.4 Bagan Kendali T^2 Hotelling

Bagan kendali T^2 Hotelling digunakan untuk mendeteksi pergeseran proses dengan menggunakan vektor mean sampel dan matriks kovariansi sampel. Nilai T^2 untuk masing-masing sampel adalah

$$T_i^2 = n(\bar{X}_{ij} - \bar{\bar{X}})' S^{-1} (\bar{X}_{ij} - \bar{\bar{X}})$$

dengan n adalah ukuran masing-masing sampel dan S^{-1} merupakan invers dari matriks kovariansi S sedangkan batas pengendali atas (BPA) dapat ditentukan dari persamaan berikut :

$$BPA = \left(\frac{mnp - mp - np + p}{mn - m - p + 1} \right) F_{\alpha, p, (mn - m - p + 1)}$$

dengan m menyatakan banyaknya sampel yang masing – masing berukuran n . Jika nilai T_i^2 untuk sampel ke- i , $T_i^2 > BPA$ maka hal ini menunjukkan bahwa sampel ke- i di luar kendali.

Jika pada saat proses pengamatan tidak terdapat subgrup atau data bersifat individu maka digunakan bagan kendali T^2 Hotelling individu. Berikut perhitungan peta kendali T^2 -Hotelling individu (Montgomery, 2009).

$$T_i^2 = (X_{ij} - \bar{X}_j)' S^{-1} (X_{ij} - \bar{X}_j) \quad i = 1, 2, \dots, n, 1, 2, \dots, p \quad (5)$$

Batas kendali T^2 Hotelling untuk pengamatan individual dengan m sampel, dapat dirumuskan dengan persamaan (2.13)s

$$BPA = \frac{(m-1)^2}{m} \beta_{\alpha, \frac{p}{2}, (m-p-1)/2} \quad (6)$$

$$BPB = 0$$

2.5 Indeks Kemampuan Proses

Analisis Kapabilitas proses adalah suatu analisis yang digunakan untuk menaksir kemampuan proses. Analisis kapabilitas proses dapat digunakan setelah proses produksi telah terkendali secara statistik. Berikut merupakan kriteria nilai kapabilitas [2].

1. $Cp = 1$, Kemampuan proses sesuai
2. $Cp > 1$, Kemampuan proses sangat baik (tingkat presisi dan akurasi tinggi)
3. $Cp < 1$, Kemampuan proses tidak baik

Rumus Indeks Kemampuan Proses (Cp) multivariat :

$$(Cp) = \frac{K}{\chi_{p, 0.95}^2} \left(\frac{(n-1)p}{s} \right)^{\frac{1}{2}} \quad (7)$$

dengan :

$$s = \sum_{i=1}^n (X_i - \bar{X}_j) A^{-1} (X_i - \bar{X}_j) \quad (8)$$

$$A^{-1} = X_i' X_i$$

$$K = \sqrt{(\bar{X}_j - \varepsilon_j)' V_0^{-1} (\bar{X}_j - \varepsilon_j)} \quad (9)$$

$$i = 1, 2, 3, \dots, n; j = 1, 2, 3$$

$$\varepsilon_j = \frac{1}{2} (BSA + BSB)$$

3. Metode Analisis

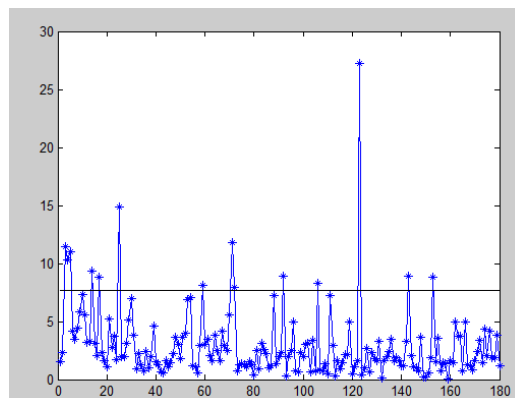
Langkah-langkah yang dilakukan berdasarkan tujuan penelitian adalah sebagai berikut :

1. Pengambilan data sekunder PT Petrokimia Gresik berupa data parameter kualitas Clay Putih sebagai material balance atau bahan pengisi dan pengikat dalam proses produksi pupuk NPK Phonska, yaitu H_2O , Mesh, dan Derajat Putih.
2. Menguji korelasi untuk mengetahui hubungan antar variabel.
3. Melakukan uji distribusi normal multivariat agar asumsi dasar yang harus dipenuhi dalam analisis multivariat terpenuhi.
4. Mengestimasi vektor rata-rata dan matriks kovariansi berdasarkan algoritma *fast-MCD*.
5. Menghitung bagan kendali T^2 Hotelling klasik dan T^2 Hotelling dengan algoritma *fast-MCD*.
6. Membandingkan nilai T^2 setiap pengamatan dengan BPA. Jika $T_i^2 \leq BPA$ maka data pengamatan ke- i berada pada batas kendali (*in control*). Sebaliknya jika $T_i^2 > BPA$, maka data pengamatan ke- i berada di luar kendali (*out of control*).
7. Menentukan kapabilitas proses dari bagan kendali T^2 Hotelling dengan menggunakan data Clay Putih sebagai material balance dalam proses produksi pupuk NPK Phonska.
8. Menarik kesimpulan dari hasil penelitian.

4. Hasil dan Pembahasan

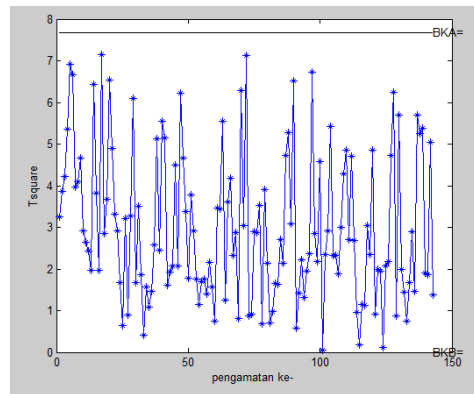
Pada penelitian ini, analisis vektor rata-rata dan matriks kovarian pada bagan kendali T^2 Hotelling dengan algoritma *Fast-MCD* akan dilakukan pada data sekunder produksi pupuk bahan Clay Putih.

Hasil perhitungan T^2 Hotelling menunjukkan bahwa terdapat 14 pengamatan yang *out of control* untuk data produksi pupuk bahan Clay Putih.



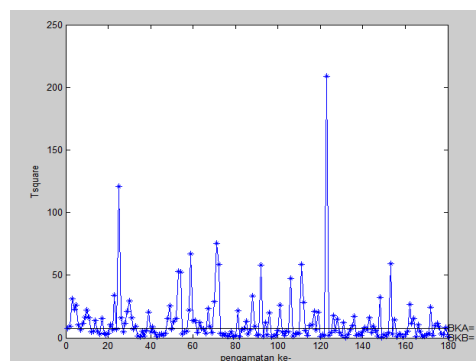
Gambar 1. bagan kendali T^2 Hotelling dengan pengamatan yang *out of control*

Untuk mencapai proses *in control*, data yang *out of control* perlu dihilangkan. Gambar 2. pada bagan kendali T^2 Hotelling menunjukkan bahwa untuk mencapai proses *in control* perlu di lakukan 9 kali iterasi penghapusan data *out of control*.



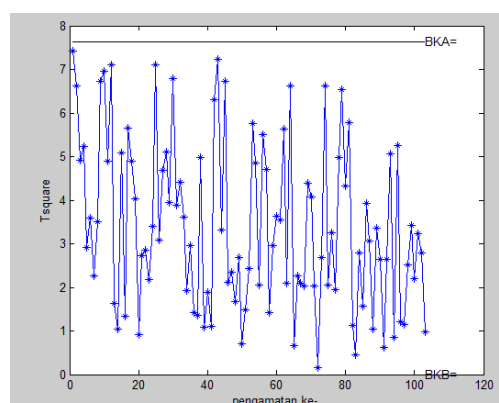
Gambar 2. bagan kendali T^2 Hotelling dengan pengamatan *in control*

Selanjutnya bagan kendali T^2 Hotelling dengan algoritma *Fast-MCD* dapat kita lihat pada Gambar 3.



Gambar 3. Bagan kendali T^2 Hotelling dengan algoritma *Fast MCD*

Gambar tersebut menunjukkan bahwa terdapat 75 nilai T_{mcd}^2 yang *out of control*. Untuk mencapai proses *in control* seperti pada Gambar 4, data yang *out of control* perlu dihilangkan. Dari hasil perhitungan diperoleh bahwa proses *in control* dapat dicapai hanya dengan 2 kali iterasi penghapusan data *out of control*.



Gambar 4. Bagan kendali T^2 Hotelling dengan algoritma *Fast-MCD*

Perbandingan hasil perhitungan bagan kendali T^2 Hotelling klasik dan bagan kendali T^2 Hotelling dengan algoritma *Fast MCD* dapat dilihat pada Tabel 1.

Tabel 1. Perbandingan hasil perhitungan T^2 dan $T_{fastMCD}^2$

Identifikasi	Data pupuk clay putih	
	T^2	$T_{fastMCD}^2$
data <i>out of control</i>	14	75
Jumlah iterasi	9	2
BPA	7,796	7,796

Dari Tabel tersebut dapat kita lihat bahwa data *out of control* lebih banyak ditemukan pada bagan kendali T^2 Hotelling dengan algoritma *Fast-MCD*. Hal ini menunjukkan bahwa bagan kendali T^2 Hotelling dengan algoritma *Fast-MCD* lebih efektif dalam mendeteksi pengamatan yang mengandung *outlier*.

Indeks Kemampuan Proses

Setelah diperoleh bagan kendali T^2 Hotelling dalam keadaan pengamatan *in control*, maka langkah selanjutnya adalah menaksir kapabilitas proses (C_p) dimana sesuai dengan batas spesifikasi yang telah ditentukan oleh perusahaan. Nilai C_p yang diperoleh untuk bagan kendali T^2 Hotelling produksi pupuk dengan bahan Clay Putih sebesar 0,1693. Sedangkan nilai C_p yang diperoleh untuk bagan kendali T^2 Hotelling dengan algoritma *Fast-MCD* yaitu 0,2707.

5. Kesimpulan

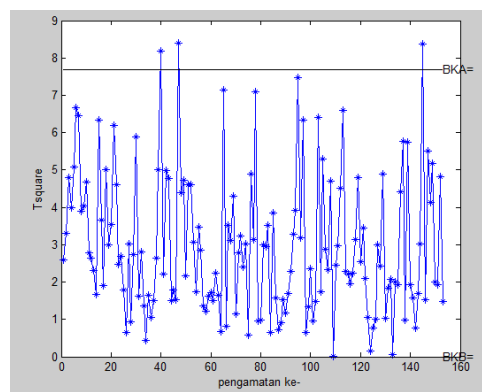
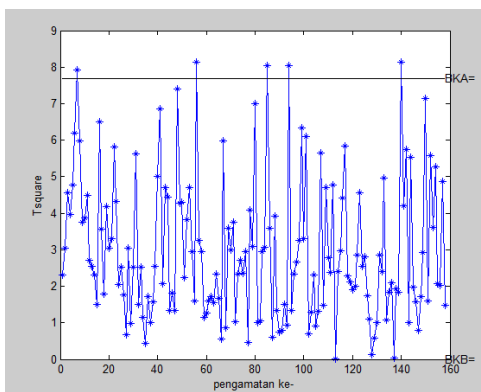
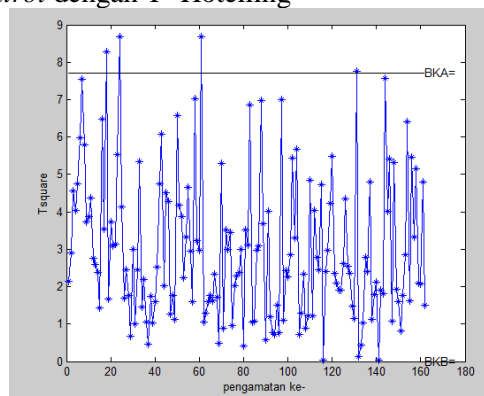
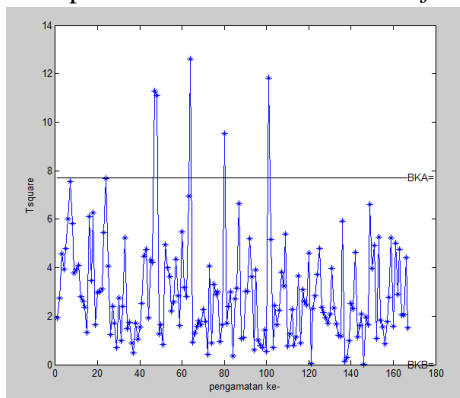
Perbandingan yang diperoleh adalah bagan kendali T^2 Hotelling dengan algoritma *Fast-MCD* lebih efektif mendeteksi data yang mengandung *outlier*. Dari hasil perhitungan diperoleh 14 data yang *out of control* pada bagan kendali T^2 Hotelling klasik dan 75 data yang *out of control* pada bagan kendali T^2 Hotelling dengan algoritma *Fast-MCD* dengan nilai BPA = 7,796.

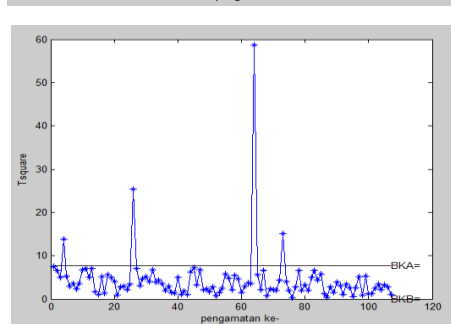
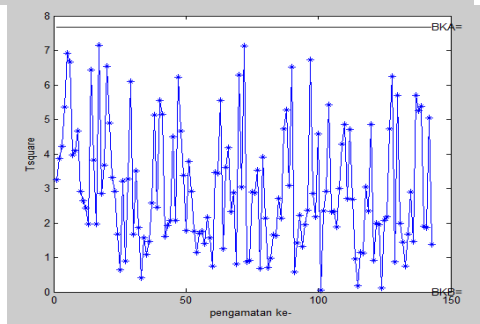
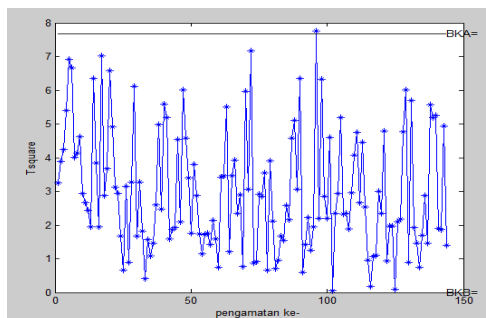
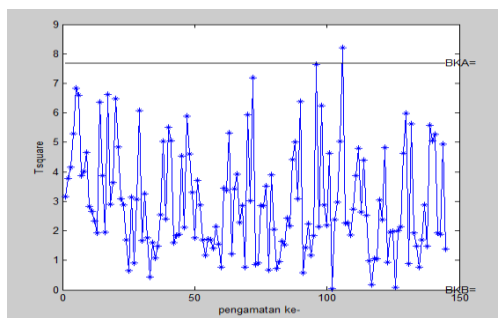
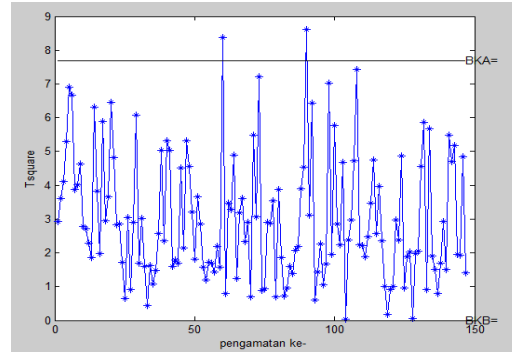
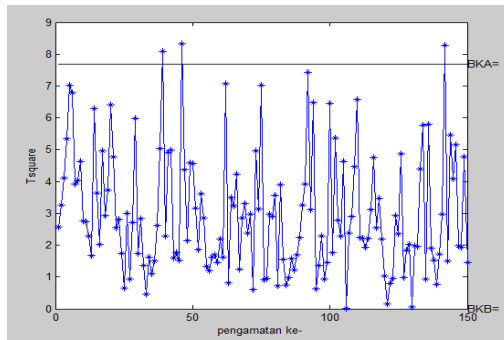
Daftar Pustaka

- [1]. Barnett V, Lewis T., 1994. *Outliers in Statistical Data*. Chichester: John Wiley & Sons, Ltd.
- [2]. Bothe, Davis,R.,1997. *Measuring Process Capability*. McGraw-Hill Companies, United States of America.
- [3]. Chenouri, S., Steiner, S. H., Variyath, A. M., 2009. A Multivariate Robust Control Chart for Individual Observations. *Journal of Quality Technology*, 41, (3).
- [4]. Erna, 2014. *Bagan Kendali T^2 Hotelling Dengan Sampel Ganda Dan Aplikasinya*. Skripsi. Program Studi Matematika Fakultas Matematika dan Ilmu pengetahuan Alam. Makassar: Universitas Hasanuddin.
- [5]. Hotelling, H., 1947. In *Techniques of Statistical Analysis*, C. Eisenhart, H. Hastay, and W. A.Wallis, eds., pp. New York, NY: McGraw-Hill.
- [6]. Hubert, M dan Driessen, 2002. Fast and Robust Discriminant Analysis. *Computational Statistics and Data Analysis*.45.
- [7]. Johnson, Richard. Dean Wichern., 2007. *Applied Multivariat Statistical Analysis, 5th ed*. New Jersey: Prentice Hall.

- [8]. Montgomery, D.C.(2009). *Introduction to statistical quality control* (6th ed.) Arizona State University: John Wiley & Sons, Inc.
- [9]. Ronal, E Walpole. 2002. *Probability and Statistics for Engineers and Scientist 7th*. Prentice Hall, Inc: Upple Saddle River, New Jersey 017458.
- [10]. Rousseeuw, P. J. (1985). *Multivariate estimation with high breakdown point*. In Mathematical Statistics and Applications B (W. Grossmann, G. Pflug, I. Vincze and W. Werz, eds.) 283297. Reidel, Dordrecht.
- [11]. Rousseeuw, P. J., Driessen, K. 1999. *A Fast Algorithm for the Minimum Covariance Determinant Estimator*. Wilrijk, Belgium: Department of Mathematics and Computer Science, University of Antwerpen.
- [12]. Todorov, V., & Filzmoser, P. *An Object Oriented Framework for Robust Multivariate Analysis*. Vienna, Austria: United Nations Industrial Development Organization (UNIDO).
- [13]. Variyath, Asokan M. dan Vattathoor, Jayasankar. 2013. Robust Control Charts for Monitoring Process Mean of Phase-I Multivariate Individual Observations. *Journal of Quality and Reliability Engineering*, Volume 2013, Article ID 542305. Hindawi Publishing Corporation
- [14]. Walpole, R.E., Myers, R.H., Myers, S.L., &Ye, K .(2012). *Probability & Statistics for Engineers & Scientists* (9th ed.). United States of America: Person Education Inc.

Lampiran : Identifikasi data *Out Of Control* dengan T^2 Hotelling





Lampiran : Identifikasi data *Out Of Control* T^2 Hotelling dengan algoritma *fast-MCD*

